



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 179453

TO: Konstantina Katcheves

Location: 2a60 / 2c70

Art Unit: 1636

Friday, February 17, 2006

Case Serial Number: 10/627007

From: Noble Jarrell

Location: Biotech-Chem Library

Rem 1B71

Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

This Page Blank (uspto)

179453

STIC-Biotech/ChemLib

From: Chan, Christina
Sent: Monday, February 13, 2006 3:47 PM
To: Katcheves, Konstantina; STIC-Biotech/ChemLib
Subject: RE: 10-627007

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----

From: Katcheves, Konstantina
Sent: Monday, February 13, 2006 3:05 PM
To: Chan, Christina
Subject: 10-627007

Christina:

Would you approve a search of SEQ ID NO:5

Thanks,
Tina

Konstantina Katcheves
Patent Examiner, AU1636
Phone: (571) 272-0768
Room: REM 2A60
Mail: REM 2C70

2/13/2006

Noted
For 2/17/05
INA
5PR Comp/prop
38ML

This Page Blank (uspto)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2006, 05:20:28 ; Search time 2087 Seconds
(without alignments)
2062.488 Million cell updates/sec

Title: US-10-627-007-5

Perfect score: 92

Sequence: 1 cggatcaggagaccctcg.....cagccgggtgctggcttttt 92

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92	100.0	625	10	CW179236
2	92	100.0	878	10	CL672134
3	76.8	83.5	775	10	CL664024
4	64	69.6	242	2	BG895158
5	47.2	51.3	1082	9	AF094862
6	39	42.4	904	10	CNS01GKM
7	30.2	32.8	853	2	BF676283
8	29.8	32.4	755	10	AG414181
9	29	31.5	976	11	CNS05RUK
10	28.8	31.3	469	9	AZ087480
11	28.8	31.3	615	6	CD728108
12	28.8	31.3	744	5	BU231483
13	28.8	31.3	746	5	BU365131
14	28.8	31.3	774	5	BU427039
15	28.8	31.3	780	6	CF256576
16	28.8	31.3	791	6	CF254599
17	28.8	31.3	804	5	BU478133
18	28.8	31.3	1272	9	BZ579155
19	28.6	31.1	345	1	AV209814
20	28.6	31.1	572	8	CX956602
21	28.6	31.1	572	8	CX956603
22	28.6	31.1	579	3	BI344684

23	28.6	31.1	681	7	CR568140
24	28.6	31.1	698	7	CR568141
25	28.6	31.1	727	8	DR878616
26	28.6	31.1	727	8	DR878617
27	28.6	31.1	779	3	BI333728
28	28.4	30.9	741	10	CG900138
29	28.2	30.7	372	4	AK203556
30	28.2	30.7	739	2	BE544496
31	28	30.4	669	1	AI322522
32	27.8	30.2	594	9	AZ624975
33	27.8	30.2	754	9	AZ894224
34	27.8	30.2	768	11	CR163471
35	27.8	30.2	888	6	CF592897
36	27.6	30.0	240	1	AA125488
37	27.6	30.0	411	1	AI596810
38	27.6	30.0	601	6	CA883090
39	27.6	30.0	638	6	CD804829
40	27.6	30.0	672	1	BB185328
41	27.6	30.0	692	5	BY728178
42	27.6	30.0	1037	8	DR128951
43	27.6	30.0	1038	4	AK021089
44	27.6	30.0	1586	4	AK039322
45	27.4	29.8	388	1	AL908917

ALIGNMENTS

RESULT 1
CW179236
LOCUS

DEFINITION

CW179236 625 bp DNA linear GSS 29-OCT-2004
104 593 11159994 116 36626 075 Sorghum methylation filtered library
(LibID: 104) Sorghum bicolor genomic clone 11159994, genomic survey
sequence.

ACCESSION
CW179236

VERSION
GSS

KEYWORDS
GSS

SOURCE
Sorghum bicolor (sorghum)

ORGANISM
Sorghum bicolor

REFERENCE
Bedell J.A., Budiman M.A., Nunberg A., Citek R.W., Robbins D.,
Jones J., Flick E., Rohlfing T., Fries J., Bradford K.,
McMenamy J., Smith M., Holsman H., Roe B.A., Wiley G., Korf I.F.,
Rabinowicz P.D., Lakey N., McCombie W.R., Jeddleloh J.A. and
Martensen R.A.

AUTHORS

TITLE

JOURNAL

PUBMED

COMMENT

15660154

Contact: Bedell JA

Orion Genomics, LLC

4041 Forest Park Ave, St. Louis, MO 63108, USA

Tel: 314 615 6979

Fax: 314 615 5975

Email: jbedell@oriongenomics.com

Plate: 593 row: f column: 18

Seq primer: T3 Reverse

Class: methylation filtered

High quality sequence stop: 625.

Location/Qualifiers

1..625

/organism="Sorghum bicolor"

/mol_type="genomic DNA"

/cultivar="ATx623"

/db_xref="taxon:4558"

/clone="11159994"

/clone_lib="Sorghum methylation filtered library (LibID: 104)"

/notes="Organ: leaf; Vector: pBCSK(-); Site: 1: HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5

kb fraction, ligated into HincII-digested pBSK(-) vector and electroporated into E. coli cells. This is a methylation filtered library."

ORIGIN
Query Match 100.0%; Score 92; DB 10; Length 625;
Best Local Similarity 100.0%; Pred. No. 4.5e-21;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 60
|||||
Db 12 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 71
|||||
QY 61 GTATTACTTAGCAGCGGGTGTGCTTTT 92
|||||
Db 72 GTATTACTTAGCAGCGGGTGTGCTTTT 103
|||||

RESULT 2

CL672134/c
LOCUS
DEFINITION
PRI016a.D04 - PRI016a.B21 (878) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CL672134 878 bp DNA linear GSS 09-JUL-2004
PRI016a.D04 - PRI016a.B21 (878) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
COMMENT

CL672134.1 GI:50172033
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 878)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satelitte organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

FEATURES

source
1..878
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match 100.0%; Score 92; DB 10; Length 878;
Best Local Similarity 100.0%; Pred. No. 4.8e-21;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 60
|||||
Db 294 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 235
|||||
QY 61 GTATTACTTAGCAGCGGGTGTGCTTTT 92
|||||
Db 234 GTATTACTTAGCAGCGGGTGTGCTTTT 203
|||||

RESULT 3

CL664024/c
LOCUS
DEFINITION

PRI0146a.A06 - PRI0146a.B21 (775) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CL664024 775 bp DNA linear GSS 09-JUL-2004
PRI0146a.A06 - PRI0146a.B21 (775) Mixed stage fosmid library of P.
pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.

CL664024.1 GI:50153989
Pristionchus pacificus
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

1 (bases 1 to 775)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
AppADB: an AcedB database for the nematode satelitte organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
14681447

Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: raif.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.
Seq primer: T7
Class: fosmid ends.

Location/Qualifiers
1..775
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

Query Match 83.5%; Score 76.8; DB 10; Length 775;
Best Local Similarity 97.5%; Pred. No. 9.8e-16;
Matches 78; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 60
|||||
Db 296 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 237
|||||
QY 61 GTATTACTTAGCAGCGGG 80
|||||
Db 236 GTATTACTTAGCAGTCCGG 217
|||||

FEATURES
source

1..775
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pEpifos-5 Fosmid vector"

ORIGIN

Query Match 83.5%; Score 76.8; DB 10; Length 775;
Best Local Similarity 97.5%; Pred. No. 9.8e-16;
Matches 78; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 60
|||||
Db 296 CGCGATCAGGAGACCTCGCGAGAACCTGAAAGCAGACATTCCTCATTGCTTCCA 237
|||||
QY 61 GTATTACTTAGCAGCGGG 80
|||||
Db 236 GTATTACTTAGCAGTCCGG 217
|||||

RESULT 4

BG895158/c
LOCUS
DEFINITION
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BG895158 242 bp mRNA linear EST 05-JUN-2001
358371 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
BG895158
BG895158.1 GI:14305399
EST.
Sus scrofa (pig)
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
1 (bases 1 to 242)
Fahrenkrug,S.C., Smith,T.P.L., Frenking,B.A., Cho,J., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Pertea,G., Sultana,R.,
Quackenbush,J. and Keele,J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
1226715

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@mail.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.

PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCCAGTCACGACG
 Plate: 121 row: L column: 2
 Seq primer: ATTAGGTGACACTATAG.

FEATURES

source

1. 242
 /organism="Sus scrofa"
 /mol_type="mRNA"
 /db_xref="taxon:9823"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /clone_lib="MARC 1P1G"
 /notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

ORIGIN

Query Match 69.6%; Score 64; DB 2; Length 242;
 Best Local Similarity 98.7%; Pred. No. 2.3e-11;
 Matches 75; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CGCGATCAGAGACCTCGCGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTT-CC 59
 |||||
 DB 76 CGCGATCAGAGACCTCGCGAGAACCTGAAGCAGCAGCATTTGCTCACATTGCTTNC 17
 |||||
 QY 60 AGTATTACTTAGCCAG 75
 |||||
 DV 16 AGTATTACTTAGCCAG 1

RESULT 5

AF094862/c
 LOCUS AF094862 Salmonella typhimurium LT2, Lambda DASH II Salmonella
 DEFINITION typhimurium genomic clone 261-17, genomic survey sequence.
 ACCESSION AF094862
 VERSION AF094862.1 GI:4322704
 KEYWORDS GSS.
 SOURCE Salmonella typhimurium
 ORGANISM Salmonella typhimurium
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 1082)
 AUTHORS Wong, R.M.Y. and McClelland, M.
 TITLE End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
 Li-Cor
 JOURNAL Unpublished (1999)
 COMMENT Contact: McClelland M
 Molecular Biology
 Sidney Kimmel Cancer Center
 3099 Science Park Road, San Diego, CA 92121, USA
 Email: mclelland@lifesci.sdsu.edu
 Class: shotgun.

FEATURES

source

1. 1082
 /organism="Salmonella typhimurium"
 /mol_type="genomic DNA"
 /strain="LT2"
 /db_xref="taxon:602"
 /clone_lib="261-T7"
 /note="Vector: Lambda DASH II; sequenced using Li-Cor
 sequencer"

ORIGIN

Query Match 51.3%; Score 47.2; DB 9; Length 1082;
 Best Local Similarity 80.9%; Pred. No. 2.4e-05;
 Matches 55; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 25 GAACCTGAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTACTTACCGCCGGTGCT 84
 |||||
 DB 1078 GAGACTGAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTATTGCCCAGCTTTGCT 1019
 |||||
 QY 85 GGCTTTT 92
 |||||
 DB 1018 GGCTTTT 1011

RESULT 6

LOCUS

DEFINITION

CNS01GKM
 Anopheles gambiae GSS SP6 end of clone 06P08 of NotreDamel library
 from strain PBST of Anopheles gambiae (African malaria mosquito),
 genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL143255
 AL143255.1 GI:7001417
 GSS.
 Anopheles gambiae (African malaria mosquito)
 Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
 Culicidae; Anophelinae; Anopheles.

REFERENCE

AUTHORS

TITLE

JOURNAL

Direct Submission
 Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

REFERENCE

AUTHORS

TITLE

JOURNAL

2 (bases 1 to 904)
 Direct Submission
 Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
 Roux, Paris 75015, France
 This clone is from an A. gambiae BAC library provided by F.H.
 Collins and sequenced by Genoscope in collaboration with the
 Laboratory of Biochem. and Biol. Molec. of Insects, Institut
 Pasteur.

FEATURES

source

1. 904
 /organism="Anopheles gambiae"
 /mol_type="genomic DNA"
 /strain="PEST"
 /db_xref="taxon:7165"
 /clone_lib="06P08"
 /clone_lib="NotreDamel"
 /note="end : SP6"

ORIGIN

Query Match 42.4%; Score 39; DB 10; Length 904;
 Best Local Similarity 100.0%; Pred. No. 0.017;
 Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 GAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTACTT 69
 |||||

DB 1 GAAGCAGCAGCATTTGCTCACATTGCTTCCAGTATTACTT 39
 |||||

RESULT 7

LOCUS

DEFINITION

BF676283/c
 BF676283
 602086942F1 NIH_MGC_83 Homo sapiens cDNA clone IMAGE:4250842 5',
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

BF676283
 BF676283.1 GI:11950178
 EST.
 Homo sapiens (human)

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1073 row: d column: 11
High quality sequence stop: 332.
FEATURES
Location/Qualifiers
1..853
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4250842"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH MGC 83"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site 1: SfII (ggcgctcgcc); Site 2: SfII
(ggcattatggcc); 5' and 3' adaptor were used in cloning
as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCGCGCGCGCCGACATG-DT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4
kb (range 0.5-4.0 kb). 14/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."
```

```

Query Match 32.8%; Score 30.2; DB 2; Length 853;
Best Local Similarity 65.7%; Pred. No. 20;
Matches 44; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 13 GACCTCGCGGAGAACCTGAAGCAGCAGATTGCTCACATTGCTTCCAGTATTACTTACG 72
DB 237 GGCCTTTGTCGAGCGGAGAGGAGGACAGTGCCACACGCTTAGGGGATTCCAGGGC 178
QY 73 CAGCCGG 79
DB 177 CAGCTGG 171

RESULT 8
AG414181
LOCUS Mus musculus molossinus DNA, clone:MSMg01-275N09.TU, genomic survey
DEFINITION sequence.
ACCESSION AG414181.1 GI:48057007
VERSION AG414181
KEYWORDS Mus musculus molossinus (Japanese wild mouse)
SOURCE Mus musculus molossinus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Abe.K., Noguchi,H., Tagawa,K., Yuzuriha,M., Toyoda,A., Kojima,T.,
Ezawa,K., Saitou,N., Hattori,M., Sakaki,Y., Moriwaki,K. and
Shiroishi,T.
TITLE Contribution of Asian mouse subspecies Mus musculus molossinus to
genomic constitution of strain C57BL/6J, as defined by BAC-end
sequence-SNP analysis
```

```

JOURNAL PUBMED
REFERENCE 2 (bases 1 to 755)
AUTHORS Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa, 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the mouse BAC library MSMg01. For BAC
library availability, please contact Kuniya Abe (abe@tc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyada, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@tc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY Vector : pBACe3.6
R.Site 1 : EcoRI.
R.Site 2 : EcoRI.
FEATURES
Location/Qualifiers
1..755
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMg01-275N09.TJ"
/sex="male"
/tissue_type="mixture of kidney and spleen"
/clone_lib="MSMg01 Mouse Male BAC Library"
```

```

Query Match 32.4%; Score 29.8; DB 10; Length 755;
Best Local Similarity 60.5%; Pred. No. 27;
Matches 49; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GGAAGACCCCTCGCGAGAACCTGAAGCAGCAGATTGCTCACATTGCTTCCAGTATTACT 68
DB 485 GGAGGGCTCTCTGAGAAGCAGGTGTGAGCAGAGCGGACATCTCTTCTTCCAGTACATCT 544
QY 69 TAGCCAGCGGGTGTGCTT 89
DB 545 TACCCAGCTGGTTCAGGCTT 565

RESULT 9
CNS05RUK/
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 039N23 of library A from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION ALJ51029
VERSION ALJ51029.1 GI:8244799
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
REFERENCE 1
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernat,A., Fizesan,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Estimating of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
PUBMED 10835645
REFERENCE 2
AUTHORS Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
```


Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
 Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
 Genome Res. 10 (7), 939-949 (2000)
 10899143
 3 (bases 1 to 976)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source
 Location/Qualifiers
 1. 976
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="039N23"
 /clone_lib="A"
 /notes="Genoscope sequence ID : COAA039CG12C1 end : T7"

ORIGIN
 Query Match 31.5%; Score 29; DB 11; Length 976;
 Best Local Similarity 63.8%; Pred. No. 55;
 Matches 44; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
 QY 22 GGAGACCTGAAGCAGCATGCTCACATGCTTCCAGTATTACTTAGCCAGCGGGT 81
 DB 684 GCAAGCTGACTACATCATCAGCGCTCAGAGCTGACGTTTGTAGCTCTCATGCTGTT 625
 QY 82 GCTGGCTTT 90
 DB 624 GTTGCTTT 616

RESULT 10
 AZ087480/c
 LOCUS
 DEFINITION
 RPCI-23-425M13.TJB RPCI-23 Mus musculus genomic clone
 RPCI-23-425M13, genomic survey sequence.
 ACCESSION
 AZ087480
 VERSION
 AZ087480.1 GI:7729528
 KEYWORDS
 GSS.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.
 1 (bases 1 to 469)
 REFERENCE
 AUTHORS
 Zhao,S., Niernan,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,B., Levins,M., Megann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)
 Other GSSs: RPCI-23-425M13.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

Plate: 425 row: M column: 13
 Seq primer: SP6
 Class: BAC ends

FEATURES
 source
 Location/Qualifiers
 1. 469
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-425M13"
 /sex="female"
 /lab_host="DH10B"
 /clone_lib="RPCI-23"
 /notes="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN
 Query Match 31.3%; Score 28.8; DB 9; Length 469;
 Best Local Similarity 62.5%; Pred. No. 54;
 Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
 QY 2 GCATCAGGAGACCTCGGAGACCTGAAGCAGCAGCATGCTCACATGCTTCCAG 61
 DB 120 GTGGTTAAGAGGACACACGCTGCTTGTAAAGACACAGTAGATCAATCATCCAG 61
 QY 62 TATTACTTAGCC 73
 DB 60 AATTCGACGCC 49

RESULT 11
 CD728108
 LOCUS
 DEFINITION
 4033546 1GAL - Chicken Intestinal Lymphocyte Gallus gallus CDNA clone 1GAL_81021 5', mRNA sequence.
 ACCESSION
 CD728108
 VERSION
 CD728108.1 GI:32278957
 KEYWORDS
 EST.
 SOURCE
 Gallus gallus (chicken)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 1 (bases 1 to 615)
 REFERENCE
 AUTHORS
 Min,W., Lillehoj,H.S., Ashwell,C.M., Matukumalli,L.K., van Tassel,C. and Han,J.Y.
 Chicken intestinal lymphocyte EST database as a resource for the analysis of mucosal immune function
 Unpublished (2003)
 JOURNAL
 COMMENT
 Contact: Hyun S. Lillehoj
 Animal Parasites Diseases Laboratory
 Animal and Natural Resources Institute, USDA
 Bldg.1043, BARC-East, Beltsville, MD 20705, USA
 Tel: 3015048771
 Fax: 3015045103
 Email: hlilleho@anri.barc.usda.gov
 Single pass sequencing. Bases called and trimmed with phred 0.000925 using options -trim_alt - -trim_fasta. Vector identified by cross match using options -minmatch 12 -minscore 18
 Plate: 81 row: O column: 21
 Seq primer: ATTAGGTGACACTAG
 High quality sequence stop: 615.

FEATURES
 source
 Location/Qualifiers
 1. 615
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="white leghorn SC"
 /db_xref="taxon:9031"

```

/clone="IGAL_81021"
/sex="mixed"
/tissue_type="Gut"
/cell_type="Lymphocyte"
/dev_stage="Adult"
/lab_host="EMDH10B"
/clone_lib="IGAL - Chicken Intestinal Lymphocyte"
/notes="Organ: Intestine; Vector: pCMV-SPORT6; Site_1:
Sali; Site_2: NotI; Normalized library from chicken gut
infected with coccidia duodenum and middle gut."

```

ORIGIN

```

Query Match      31.3%; Score 28.8; DB 6; Length 615;
Best Local Similarity 58.0%; Pred. No. 58;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY  4 GATCAGGAGACCTCGCGAGAACCTGAAAGCAGCAGACATTGCTCACATTGCTTCCAGTA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  30 GGTTCGAATGACCTCTGCTGAGAACATGCTACCTCTGTTTGTGTCGACATTGTCGATGC 89

QY  64 TTACTTAGCCAGCCGGTCTGCTTTT 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  90 TTATCATTCGAACTGGTCTGCTTTT 117

```

RESULT 12

```

BU231483
LOCUS
DEFINITION
  BU231483 744 bp mRNA linear EST 26-NOV-2002
  603948804F1 CSEQCHN23 Gallus gallus cDNA clone CHEST905h1 5', mRNA
  sequence.
ACCESSION
  BU231483
VERSION
  BU231483.1 GI:25473276
KEYWORDS
  EST.
SOURCE
  Gallus gallus (chicken)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
  1 (bases 1 to 744)
  Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
  Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
  A Comprehensive Collection of Chicken cDNAs
  Curr. Biol. 12 (22), 1965-1969 (2002)
  12445392
  PUMED
  COMMENT
    Contact: Simon Hubbard
    Department of Biomolecular Sciences
    University of Manchester Institute of Science and Technology
    (UMIST)
    PO Box 88, Manchester, M60 1QD, UK
    Tel: 01612008930
    Fax: 01612360409
    Email: Simon.Hubbard@umist.ac.uk.

```

FEATURES

```

Location/Qualifiers
1..744
  /organism="Gallus gallus"
  /mol_type="mRNA"
  /strain="White Leghorn, Hixex"
  /db_xref="taxon:9031"
  /clone="CHEST905h1"
  /dev_stage="22"
  /lab_host="DH10B"
  /clone_lib="CSEQCHN23"
  /note="Organ: Head; Vector: pBluescript II KS(+); Site_1:
  EcoRI; Site_2: NotI; This normalized library was
  constructed from 1 million independent clones. cDNA
  synthesis was initiated using an oligo(dT) primer, using
  methylated C in the first strand synthesis reaction.
  Following this first strand reaction, double-stranded cDNA
  was blunted, ligated to NotI adaptors, digested with
  EcoRI, size-selected, and cloned into the NotI and EcoRI
  compatible sites of a custom modified MCS of the
  pBluescript (KS+) vector. The library was normalized in 2
  rounds using conditions adapted from Soares et al., PNAS

```

```

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

```

Query Match      31.3%; Score 28.8; DB 5; Length 744;
Best Local Similarity 58.0%; Pred. No. 61;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY  4 GATCAGGAGACCTCGCGAGAACCTGAAAGCAGCAGACATTGCTCACATTGCTTCCAGTA 63
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  107 GGTTCGAATGACCTCTGCTGAGAACATGCTACCTCTGTTTGTGTCGACATTGTCGATGC 166

QY  64 TTACTTAGCCAGCCGGTCTGCTTTT 91
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  167 TTATCATTCGAACTGGTCTGCTTTT 194

```

RESULT 13

```

BU365131
LOCUS
DEFINITION
  BU365131 746 bp mRNA linear EST 28-NOV-2002
  603586589F1 CSEQCHN72 Gallus gallus cDNA clone CHEST543b24 5', mRNA
  sequence.
ACCESSION
  BU365131
VERSION
  BU365131.1 GI:25873132
KEYWORDS
  EST.
SOURCE
  Gallus gallus (chicken)
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
  1 (bases 1 to 746)
  Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
  Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
  A Comprehensive Collection of Chicken cDNAs
  Curr. Biol. 12 (22), 1965-1969 (2002)
  12445392
  PUMED
  COMMENT
    Contact: Simon Hubbard
    Department of Biomolecular Sciences
    University of Manchester Institute of Science and Technology
    (UMIST)
    PO Box 88, Manchester, M60 1QD, UK
    Tel: 01612008930
    Fax: 01612360409
    Email: Simon.Hubbard@umist.ac.uk.

```

FEATURES

```

Location/Qualifiers
1..746
  /organism="Gallus gallus"
  /mol_type="mRNA"
  /strain="Compton line 151"
  /db_xref="taxon:9031"
  /clone="CHEST543b24"
  /sex="Female"
  /tissue_type="cerebrum"
  /dev_stage="adult"
  /lab_host="DH10B"
  /clone_lib="CSEQCHN72"
  /note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
  EcoRI; Site_2: NotI; This normalized library was
  constructed from 1 million independent clones. cDNA
  synthesis was initiated using an oligo(dT) primer, using
  methylated C in the first strand synthesis reaction.
  Following this first strand reaction, double-stranded cDNA
  was blunted, ligated to NotI adaptors, digested with
  EcoRI, size-selected, and cloned into the NotI and EcoRI
  compatible sites of a custom modified MCS of the
  pBluescript (KS+) vector. The library was normalized in 2
  rounds using conditions adapted from Soares et al., PNAS
  (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
  (1996): 791, except that a significantly longer
  reannealing hybridization was used."

```

ORIGIN

Query Match 31.3%; Score 28.8; DB 5; Length 746;
 Best Local Similarity 58.0%; Pred. No. 61;
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGAGAACCTGAAGACGACATTGCTCATTGCTTCCAGTA 63
 DB 435 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 494
 QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91
 DB 495 TTATCATTTGCACTGGTGTGCTTTT 522

RESULT 14
 BU427039 774 bp mRNA linear EST 29-NOV-2002
 LOCUS 603954645F1 CSEQRBN09 Gallus gallus cDNA clone CHEST920112 5', mRNA
 DEFINITION sequence.

ACCESSION BU427039
 VERSION BU427039.1 GI:25919715
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 774)
 Boardman,P.E., Sanz-Esquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
 Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
 A Comprehensive Collection of Chicken cDNAs
 CURR. BIOL. 12 (22), 1965-1969 (2002)

12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..774
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="layer and broiler"
 /db_xref="taxon:9031"
 /clone="CHEST920112"
 /sex="Male and female"
 /tissue_type="Chondrocytes isolated from growth plate cartilage"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="CSEQRBN09"

Site 2: Not1: This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 31.3%; Score 28.8; DB 5; Length 774;
 Best Local Similarity 58.0%; Pred. No. 61;
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGAGAACCTGAAGACGACATTGCTCATTGCTTCCAGTA 63
 DB 124 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 183
 QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91
 DB 184 TTATCATTTGCACTGGTGTGCTTTT 211

RESULT 15
 CF256576 780 bp mRNA linear EST 07-AUG-2003
 LOCUS mdvnl41_d06 Marek's disease virus-infected spleen Gallus gallus
 DEFINITION cDNA, mRNA sequence.

ACCESSION CF256576
 VERSION CF256576.1 GI:33489831
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 780)
 Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,
 Chausse,A.M. and Zoorob,R.
 A collection of chicken ESTs from activated immune cells
 Unpublished (2003)

JOURNAL JOURNAL
 COMMENT Contact: Zoorob R
 UPB 1983
 CNRS
 7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
 Tel: 33 1 49 58 35 00
 Fax: 33 1 49 58 33 81
 Email: zoorob@vjf.cnrs.fr.

FEATURES
 source
 1..780
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /db_xref="taxon:9031"
 /clone_lib="Marek's disease virus-infected spleen"
 /notes="Organ: Spleen; Vector: pTriplex2"

ORIGIN
 Query Match 31.3%; Score 28.8; DB 6; Length 780;
 Best Local Similarity 58.0%; Pred. No. 61;
 Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 4 GATCAGGAAGACCTCGCGAGAACCTGAAGACGACATTGCTCATTGCTTCCAGTA 63
 DB 631 GGTTTGAATGACCTGCTGGAGAACATGCTACCTCTGTTTGTGACATTGCTGCATGC 690
 QY 64 TTACTTAGCCAGCGGCTGCTGCTTTT 91
 DB 691 TTATCATTTGCACTGGTGTGCTTTT 718

Search completed: February 16, 2006, 06:25:17
 Job time : 2091 secs

Blank page (not for
insertion)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2006, 04:56:28 ; Search time 338 Seconds
(without alignments)
1814.058 Million cell updates/sec

Title: US-10-627-007-5

Perfect score: 92

Sequence: 1 cgcgcaggaagaccctcg.....cagccgggtgctggcttttt 92

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21.*
1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003gs.*
11: Geneseqn2003ds.*
12: Geneseqn2004as.*
13: Geneseqn2004bs.*
14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	100.0	92	6 ABS63103	Abse63103 Identific
2	45.2	49.1	60203	10 ACP65383_4	Continuation (5 of
3	45.2	49.1	110000	10 ACP65385_4	Continuation (5 of
4	45.2	49.1	110000	10 ACP67367_04	Continuation (5 of
5	45.2	49.1	110000	10 ACP67367_33	Continuation (34 o
6	45.2	49.1	110000	10 ACP67367_34	Continuation (35 o
7	31	33.7	354	11 ACH94896	Ach94896 Klebsiell
8	30	32.6	9603	2 AAX60894	Aax60894 DHPR-TS g
9	28.2	30.7	1338	12 AD000018	Ado00018 Novel hum
10	28.2	30.7	1338	12 ADN98449	Adn98449 Novel hum
11	28.2	30.7	2523	5 AAS72985	Aas72985 DNA encod
12	28.2	30.7	11872	4 AAK84027	Aak84027 Human imm
13	28.2	30.7	11872	4 AAK66893	Aak66893 Human imm
14	27.6	30.0	1404	8 ACA43893	Aca43893 Prokaryot
15	27.4	29.8	8668	13 ADR84515	Adr84515 Aspergill
16	27	29.3	2337	6 ABN95768	Abn95768 Arabidops
17	27	29.3	4669	4 ABL02136	Ab102136 Drosophil
18	26.8	29.1	558	4 AAI19231	Aai19231 Probe #91
19	26.8	29.1	558	4 ABA64237	Ab64237 Human foe

ALIGNMENTS

RESULT 1

ABS63103

ID ABS63103 standard; DNA; 92 BP.

XX ABS63103;

XX 05-NOV-2002 (first entry)

DE Identification of small RNA molecules related polynucleotide #5.

XX Small RNA; sRNA; cell regulation; intercell regulation; diagnostic;

KW antibiotic; ds.

XX Escherichia coli.

XX WO200260914-A2.

XX 08-AUG-2002.

XX 31-JAN-2002; 2002WO-US003147.

XX 01-FEB-2001; 2001US-0266402P.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Gottesman S, Storz G, Repoila F, Wassarman K, Rosenow C;

XX WPI; 2002-619223/66.

XX New polynucleotide and polypeptide from small RNA candidates and open reading frames, useful as mediators of cell or intercell regulation and in the development of diagnostics and antibiotics.

XX Claim 6; Page 59; 90pp; English.

XX The invention describes an isolated Escherichia coli polynucleotide from small RNA (sRNA) candidates and open reading frames (ORFs), its complement, or its homologue. The polynucleotides or the polypeptides are useful as mediators of cell or intercell regulation. They may also be used in the development of diagnostics and antibiotics. This sequence represents a candidate polynucleotide studied for the presence of a small RNA open reading frame

Aai44394 Probe #13
Aba31380 Probe #98
Aak38436 Human bon
Aak12715 Human bra
Aab14495 Human gen
Acc49520 Tumour-as
Adq96373 T cell ac
Adq96375 T cell ac
Adq96377 T cell ac
Abz44817 Human GPC
Adq96379 T cell ac
Adb67668 Human xen
Acn37557 Tumour-as
Acc49519 Tumour-as
Aad17361 Human bon
Ade4082 Human pro
Aeb96535 Human CAB
Abl08715 Drosophil
Abl08714 Drosophil
Abk27555 DNA encod
Aal05359 Human rep
Abl98228 Human tes
Adf65864 T-cell re
Ada02957 Mouse Dad
Adb72695 Mouse Dad
Adc85437 Mouse Dad

20 26.8 29.1 558 4 AAI44394
21 26.8 29.1 558 4 ABA31380
22 26.8 29.1 558 4 AAK38436
23 26.8 29.1 558 4 AAK12715
24 26.8 29.1 558 6 AAB14495
c 25 26.8 29.1 2203 8 ACC49520
c 26 26.8 29.1 2708 12 ADQ96373
c 27 26.8 29.1 2708 12 ADQ96375
c 28 26.8 29.1 2879 12 ADQ96377
c 29 26.8 29.1 2970 8 ABZ44817
c 30 26.8 29.1 4531 12 ADQ96379
c 31 26.8 29.1 4546 9 ADB67668
c 32 26.8 29.1 4565 13 ACN37557
c 33 26.8 29.1 4577 8 ACC49519
c 34 26.8 29.1 4616 4 AAD17361
c 35 26.8 29.1 4637 10 ADE54082
c 36 26.8 29.1 151909 14 AEB96535
c 37 26.6 28.9 3750 4 ABL08715
c 38 26.6 28.9 12225 4 ABL08714
c 39 26.4 28.7 1559 6 ABK27555
c 40 26.4 28.7 16062 4 AAL05359
c 41 26.4 28.7 16062 4 ABL98228
c 42 26.2 28.5 12035 10 ADF65864
c 43 26.2 28.5 38596 9 ADA02957
c 44 26.2 28.5 38596 10 ADB72695
c 45 26.2 28.5 38596 10 ADC85437

XX	Sequence 92 BP; 21 A; 26 C; 23 G; 22 T; 0 U; 0 Other;	
SQ	Query Match Best Local Similarity 100.0%; Score 92; DB 6; Length 92; Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 CGCGATCAGGAGACCTCGCGAGAACCTGAAGCAGCATTGCTCATTGCTTCCA 60 	
Db	1 CGCGATCAGGAGACCTCGCGAGAACCTGAAGCAGCATTGCTCATTGCTTCCA 60 	
QY	61 GTATTACTTAGCCAGCCGGGCTGCTTTT 92 	
Db	61 GTATTACTTAGCCAGCCGGGCTGCTTTT 92 	
RESULT 2		
ACF65383_4	Continuation (5 of 5) of ACF65383 from base 400001 (Photorhabdus luminescens nucleotide WP Sequence split into 5 fragments LOCUS ACF65383 Accession ACF65383	
WP	Fragment Name Begin End	
WP	ACF65383_0 1 110000	
WP	ACF65383_1 100001 210000	
WP	ACF65383_2 200001 310000	
WP	ACF65383_3 300001 410000	
WP	ACF65383_4 400001 460203	
Query Match	49.1%; Score 45.2; DB 10; Length 60203;	
Best Local Similarity	93.5%; Pred. No. 4e-06;	
Matches	58; Conservative 0; Mismatches 3; Indels 1; Gaps 1;	
QY	32 AAAGCAGCATTGCTCATTGCTTCCAGTATTACTTAGCCAGC-CGGGTGCTGCTTT 90 	
Db	14503 AAAGCAGCATTGCTCATTGCTTCCAGTATTTTAAAGCAGCTCGGGTCTGCTTT 14562 	
QY	91 TT 92 	
Db	14563 TT 14564 	
RESULT 3		
ACF65385_4	Continuation (5 of 7) of ACF65385 from base 400001 (Photorhabdus luminescens nucleotide WP Sequence split into 7 fragments LOCUS ACF65385 Accession ACF65385	
WP	Fragment Name Begin End	
WP	ACF65385_0 1 110000	
WP	ACF65385_1 100001 210000	
WP	ACF65385_2 200001 310000	
WP	ACF65385_3 300001 410000	
WP	ACF65385_4 400001 510000	
WP	ACF65385_5 500001 610000	
WP	ACF65385_6 600001 618776	
Query Match	49.1%; Score 45.2; DB 10; Length 110000;	
Best Local Similarity	81.2%; Pred. No. 5e-06;	
Matches	65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;	
QY	15 CCCTCGCGAGAACCTGAAGCAGCATTGCTCATTGCTTCCAGTATTAC--TTAGC 72 	
Db	99120 CCTTCGGGTACAGATGAAGCAGCATTGCTCATTGCTTCCAGTATTACTTTAGC 99179 	
QY	73 CAGCCGGGTGCTGCTTTT 92 	
Db	99180 CAGCCAAAGTCTGCTTCTT 99199 	
RESULT 4		
ACF67367_04	Continuation (5 of 57) of ACF67367 from base 400001 (Photorhabdus luminescens nucleotide WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367	
WP	Fragment Name Begin End	
WP	ACF67367_00 1 110000	
WP	ACF67367_01 100001 210000	

WP	ACF67367_02	200001	310000	
WP	ACF67367_03	300001	410000	
WP	ACF67367_04	400001	510000	
WP	ACF67367_05	500001	610000	
WP	ACF67367_06	600001	710000	
WP	ACF67367_07	700001	810000	
WP	ACF67367_08	800001	910000	
WP	ACF67367_09	900001	1010000	
WP	ACF67367_10	1000001	1110000	
WP	ACF67367_11	1100001	1210000	
WP	ACF67367_12	1200001	1310000	
WP	ACF67367_13	1300001	1410000	
WP	ACF67367_14	1400001	1510000	
WP	ACF67367_15	1500001	1610000	
WP	ACF67367_16	1600001	1710000	
WP	ACF67367_17	1700001	1810000	
WP	ACF67367_18	1800001	1910000	
WP	ACF67367_19	1900001	2010000	
WP	ACF67367_20	2000001	2110000	
WP	ACF67367_21	2100001	2210000	
WP	ACF67367_22	2200001	2310000	
WP	ACF67367_23	2300001	2410000	
WP	ACF67367_24	2400001	2510000	
WP	ACF67367_25	2500001	2610000	
WP	ACF67367_26	2600001	2710000	
WP	ACF67367_27	2700001	2810000	
WP	ACF67367_28	2800001	2910000	
WP	ACF67367_29	2900001	3010000	
WP	ACF67367_30	3000001	3110000	
WP	ACF67367_31	3100001	3210000	
WP	ACF67367_32	3200001	3310000	
WP	ACF67367_33	3300001	3410000	
WP	ACF67367_34	3400001	3510000	
WP	ACF67367_35	3500001	3610000	
WP	ACF67367_36	3600001	3710000	
WP	ACF67367_37	3700001	3810000	
WP	ACF67367_38	3800001	3910000	
WP	ACF67367_39	3900001	4010000	
WP	ACF67367_40	4000001	4110000	
WP	ACF67367_41	4100001	4210000	
WP	ACF67367_42	4200001	4310000	
WP	ACF67367_43	4300001	4410000	
WP	ACF67367_44	4400001	4510000	
WP	ACF67367_45	4500001	4610000	
WP	ACF67367_46	4600001	4710000	
WP	ACF67367_47	4700001	4810000	
WP	ACF67367_48	4800001	4910000	
WP	ACF67367_49	4900001	5010000	
WP	ACF67367_50	5000001	5110000	
WP	ACF67367_51	5100001	5210000	
WP	ACF67367_52	5200001	5310000	
WP	ACF67367_53	5300001	5410000	
WP	ACF67367_54	5400001	5510000	
WP	ACF67367_55	5500001	5610000	
WP	ACF67367_56	5600001	5648894	
Query Match				49.1%; Score 45.2; DB 10; Length 110000;
Best Local Similarity				93.5%; Pred No. 5e-06;
Matches				58; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
Qy	32 AAAGCAGCATTGCTCATTGCTTCCAGTATTACTTAGCCAGC-CGGGTGCTGGCTTT 90			
Db	81981 AAAGCAGCATTGCTCATTGCTTCCAGTGTCTTTTAGCCAGCTCGGGTGTGGCTTT 82040			
Qy	91 TT 92			
Db	82041 TT 82042			
RESULT 5				
ACF67367_33				
Continuation (34 of 57) of ACF67367 from base 3300001 (Photorhabdus luminescens nucleotide				
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367				

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match 49.1%; Score 45.2; DB 10; Length 110000;
 Best Local Similarity 81.2%; Pred. No. 5e-06;
 Matches 65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY	15	CCCTCGCGGAGACCTGAAAGCAGCATTGCTCATTGCTTCCAGTATTAC--TTAGC 72
DB	104254	CCTTCGGGTACAGATGAAAGCAGCATTGCTCATTGCTTCCAGTGTACTTTAGC 104313
QY	73	CAGCCGGGTGCTGGCTTTT 92
DB	104314	CAGCAAGTGTGGCTTCTT 104333

RESULT 6

WP	Fragment Name	Begin	End
WP	ACF67367_00	1	110000
WP	ACF67367_01	100001	210000
WP	ACF67367_02	200001	310000
WP	ACF67367_03	300001	410000
WP	ACF67367_04	400001	510000
WP	ACF67367_05	500001	610000
WP	ACF67367_06	600001	710000
WP	ACF67367_07	700001	810000
WP	ACF67367_08	800001	910000
WP	ACF67367_09	900001	1010000
WP	ACF67367_10	1000001	1110000
WP	ACF67367_11	1100001	1210000
WP	ACF67367_12	1200001	1310000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Query Match 49.1%; Score 45.2; DB 10; Length 110000;
 Best Local Similarity 81.2%; Pred. No. 5e-06;
 Matches 65; Conservative 0; Mismatches 13; Indels 2; Gaps 1;

QY	15	CCCTCGCGGAGACCTGAAAGCAGCATTGCTCATTGCTTCCAGTATTAC--TTAGC 72
DB	4254	CCTTCGGGTACAGATGAAAGCAGCATTGCTCATTGCTTCCAGTGTACTTTAGC 4313
QY	73	CAGCCGGGTGCTGGCTTTT 92
DB	4314	CAGCAAGTGTGGCTTCTT 4333

PR 15-NOV-2002; 2002US-0426430P.
 PR 15-NOV-2002; 2002US-0426916P.
 PR 27-NOV-2002; 2002US-0429224P.
 PR 27-NOV-2002; 2002US-0429275P.
 PR 27-NOV-2002; 2002US-0429302P.
 PR 27-NOV-2002; 2002US-0429326P.
 PR 27-NOV-2002; 2002US-0429651P.
 PR 04-DEC-2002; 2002US-0430645P.
 PR 04-DEC-2002; 2002US-0430651P.
 PR 04-DEC-2002; 2002US-0430657P.
 PR 04-DEC-2002; 2002US-0430663P.
 PR 04-DEC-2002; 2002US-0430668P.
 PR 04-DEC-2002; 2002US-0430684P.
 PR 05-DEC-2002; 2002US-0430937P.
 PR 05-DEC-2002; 2002US-0430965P.
 PR 05-DEC-2002; 2002US-0431458P.
 PR 12-DEC-2002; 2002US-0433251P.
 PR 12-DEC-2002; 2002US-0433500P.
 PR 13-DEC-2002; 2002US-0433316P.
 PR 13-DEC-2002; 2002US-0433318P.
 PR 23-DEC-2002; 2002US-0436236P.
 PR 03-JAN-2003; 2003US-0437914P.
 PR 17-JAN-2003; 2003US-0440820P.
 PR 17-JAN-2003; 2003US-0440821P.
 PR 18-APR-2003; 2003US-0463700P.
 PR 18-APR-2003; 2003US-0463708P.
 PR 18-APR-2003; 2003US-0463716P.
 PR 02-MAY-2003; 2003US-0463732P.
 PR 02-MAY-2003; 2003US-0467199P.
 PR 02-MAY-2003; 2003US-0467201P.
 PR 02-MAY-2003; 2003US-0467203P.
 PR 19-MAY-2003; 2003US-0467230P.
 PR 19-MAY-2003; 2003US-0471306P.
 PR 22-MAY-2003; 2003US-0471336P.
 PR 22-MAY-2003; 2003US-0472420P.
 PR 22-MAY-2003; 2003US-0472430P.
 PR 09-JUN-2003; 2003US-0476609P.
 PR 09-JUN-2003; 2003US-0476621P.
 PR 09-JUN-2003; 2003US-0476632P.
 PR 09-JUN-2003; 2003US-0476641P.
 PR 08-JUL-2003; 2003US-0485217P.
 PR 08-JUL-2003; 2003US-0485218P.
 PR 08-JUL-2003; 2003US-0485223P.
 PR 08-JUL-2003; 2003US-0485224P.
 PR 08-JUL-2003; 2003US-0485325P.
 PR 08-JUL-2003; 2003US-0485359P.
 PR 14-JUL-2003; 2003US-0486446P.
 PR 14-JUL-2003; 2003US-0486480P.
 PR 15-JUL-2003; 2003US-0486891P.
 PR 15-JUL-2003; 2003US-0486960P.
 PR 08-AUG-2003; 2003US-0493341P.
 PR 08-AUG-2003; 2003US-0493370P.
 PR 08-AUG-2003; 2003US-0493573P.
 PR 08-AUG-2003; 2003US-0493577P.
 PA (FIVE-) FIVE PRIME THERAPEUTICS INC.
 XX
 XX Williams LT, Chu K, Lee E, Hestir K, Beaurang PA, Behrens D;
 PI Halenbeck RP, Kochakota S, Lin H, Linnemann T, Pierce K, Wang Y;
 PI Wong JGP, Wu G, Zhang H, Zeng C;
 XX
 XX WPI; 2004-365511/34.
 DR P-PSDB; AUN99233.
 XX
 XX New nucleic acid molecules, useful in preparing a composition for
 PT treating or preventing e.g. inflammatory CNS, bacterial or viral
 PT disorders, cancer, psoriasis, diabetes, ischemic heart disease or
 PT ulcerative colitis.
 XX
 XX Claim 1; SEQ ID NO 49; 532pp; English.
 PS
 XX The invention relates to a nucleic acid molecule comprising a
 CC polynucleotide sequence or its complement that encodes a polypeptide. The

CC nucleic acid is useful in preparing a composition for treating or
 CC preventing inflammatory, CNS, immune, bacterial or viral disorder,
 CC cancer, psoriasis, diabetes, early aging, hormonal imbalance, ischemic
 CC heart disease or ulcerative colitis. This sequence corresponds to a
 CC nucleic acid of the invention.
 XX
 SQ Sequence 1338 BP; 248 A; 476 C; 345 G; 269 T; 0 U; 0 Other;
 Query Match 30.7%; Score 28.2; DB 12; Length 1338;
 Best Local Similarity 59.3%; Pred. No. 3.7;
 Matches 48; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 10 GAAGACCTCGCGGAGACCTGAGAGCAGCATTGCTCACATTGCTTCAGATTACTT 69
 DB 162 GCACGCCCTTTCGCGAGCGCTGAGCCAGAGCGTCTCACCTTCCTCCACCCACAC 221
 QY 70 AGCCAGCCGCGGTCTGGCTTT 90
 DB 222 CACGAGCCGCTGGGGAGTT 242
 RESULT 11
 AAS72985
 ID AAS72985 standard; cDNA; 2523 BP.
 XX AAS72985;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #8789.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US008631.
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG08798.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 8789; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations

PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX
 XX Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-483426/52.

XX
 PT Nucleic acids encoding human immune/haematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.

XX
 PS Disclosure; SEQ ID NO 38839; 3071pp + Sequence Listing; English.

XX
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention

XX
 SQ Sequence 11872 BP; 3363 A; 2615 C; 2631 G; 3263 T; 0 U; 0 Other;

Query Match 30.7%; Score 28.2; DB 4; Length 11872;
 Best Local Similarity 57.3%; Pred. No. 8.2;
 Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 2 GCGATCAGGAGACCTCGCGGAGAACCTGAAGACGACGATTCACATTCGCTTCCAG 61
 DB 4804 GAGATCAGGACCATCTCGCTAACCGTGGAACCCCGCTTCTTAAATAACAAAAA 4863

QY 62 TATTACTTAGCCAGCCGGTGCTGCTTT 90
 DB 4864 TATAGCTGAGCATGGTGGCAGGTGCCTAT 4892

RESULT 13

AAK6893
 ID AAK6893 standard; DNA; 11872 BP.

XX AAK6893;

DT 06-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21705.
 XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

PF 17-JAN-2001; 2001WO-US001354.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226682P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.

PR 01-SEP-2000; 2000US-0229344P.

PR 01-SEP-2000; 2000US-0229345P.

PR 05-SEP-2000; 2000US-0229509P.

PR 05-SEP-2000; 2000US-0229513P.

PR 06-SEP-2000; 2000US-0230437P.

PR 06-SEP-2000; 2000US-0230438P.

PR 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.

```

PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-024617P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246612P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.

PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
XX Disclosure; SEQ ID NO 21705; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK5950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
XX Sequence 11872 BP; 3363 A; 2615 C; 2631 G; 3263 T; 0 U; 0 Other;
SQ
Query Match 30.7%; Score 28.2; DB 4; Length 11872;
Best Local Similarity 57.3%; Pred. No. 8.2;
Matches 51; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 2 GCGATCAGGAAGACCTCGGGAGAACCTGAAAGCAGCAGCATTCGTTCACATTCCTCCAG 61
Db 4804 GAGATCAAGACCATCTCTGCTAACACGGTGAACCCGCTTTACTAAATAACAAAAA 4863
QY 62 TATTACTTAGCCAGCGCGGTGTGGCTTT 90
Db 4864 TATAGCTGAGCATGGTGGCAGGTGCCTAT 4892
RESULT 14
ACA43893
ID ACA43893 standard; DNA; 1404 BP.
XX
XX ACA43893;
AC
XX
XX 19-JUN-2003 (first entry)
DT
XX
XX Prokaryotic essential gene #25550.
DE
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW

```


Query Match 100.0%; Score 92; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.3e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 60
Db 67494 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 67435

Qy 61 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 92
Db 67434 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 67403

RESULT 5
AE005674_35/c
WPCOMMENT

Sequence split into 46 fragments LOCUS AE005674 Accession AE005674

Fragment Name	Begin	End
AE005674_00	1	110000
AE005674_01	100001	210000
AE005674_02	200001	310000
AE005674_03	300001	410000
AE005674_04	400001	510000
AE005674_05	500001	610000
AE005674_06	600001	710000
AE005674_07	700001	810000
AE005674_08	800001	910000
AE005674_09	900001	1010000
AE005674_10	1000001	1110000
AE005674_11	1100001	1210000
AE005674_12	1200001	1310000
AE005674_13	1300001	1410000
AE005674_14	1400001	1510000
AE005674_15	1500001	1610000
AE005674_16	1600001	1710000
AE005674_17	1700001	1810000
AE005674_18	1800001	1910000
AE005674_19	1900001	2010000
AE005674_20	2000001	2110000
AE005674_21	2100001	2210000
AE005674_22	2200001	2310000
AE005674_23	2300001	2410000
AE005674_24	2400001	2510000
AE005674_25	2500001	2610000
AE005674_26	2600001	2710000
AE005674_27	2700001	2810000
AE005674_28	2800001	2910000
AE005674_29	2900001	3010000
AE005674_30	3000001	3110000
AE005674_31	3100001	3210000
AE005674_32	3200001	3310000
AE005674_33	3300001	3410000
AE005674_34	3400001	3510000
AE005674_35	3500001	3610000
AE005674_36	3600001	3710000
AE005674_37	3700001	3810000
AE005674_38	3800001	3910000
AE005674_39	3900001	4010000
AE005674_40	4000001	4110000
AE005674_41	4100001	4210000
AE005674_42	4200001	4310000
AE005674_43	4300001	4410000
AE005674_44	4400001	4510000
AE005674_45	4500001	4607203

Continuation (36 of 46) of AE005674 from base 3500001 (AE005674 Shigella flexneri 2a str)

Query Match 100.0%; Score 92; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.3e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 60
Db 53665 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 53606

Qy 61 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 92
Db 53605 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 53574

RESULT 6
U00096_35/c
WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096

Fragment Name	Begin	End
U00096_00	1	110000
U00096_01	100001	210000
U00096_02	200001	310000
U00096_03	300001	410000
U00096_04	400001	510000
U00096_05	500001	610000
U00096_06	600001	710000
U00096_07	700001	810000
U00096_08	800001	910000
U00096_09	900001	1010000
U00096_10	1000001	1110000
U00096_11	1100001	1210000
U00096_12	1200001	1310000
U00096_13	1300001	1410000
U00096_14	1400001	1510000
U00096_15	1500001	1610000
U00096_16	1600001	1710000
U00096_17	1700001	1810000
U00096_18	1800001	1910000
U00096_19	1900001	2010000
U00096_20	2000001	2110000
U00096_21	2100001	2210000
U00096_22	2200001	2310000
U00096_23	2300001	2410000
U00096_24	2400001	2510000
U00096_25	2500001	2610000
U00096_26	2600001	2710000
U00096_27	2700001	2810000
U00096_28	2800001	2910000
U00096_29	2900001	3010000
U00096_30	3000001	3110000
U00096_31	3100001	3210000
U00096_32	3200001	3310000
U00096_33	3300001	3410000
U00096_34	3400001	3510000
U00096_35	3500001	3610000
U00096_36	3600001	3710000
U00096_37	3700001	3810000
U00096_38	3800001	3910000
U00096_39	3900001	4010000
U00096_40	4000001	4110000
U00096_41	4100001	4210000
U00096_42	4200001	4310000
U00096_43	4300001	4410000
U00096_44	4400001	4510000
U00096_45	4500001	4610000
U00096_46	4600001	4639675

Continuation (36 of 47) of U00096 from base 3500001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 92; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.3e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 60
Db 79040 CGCGATCAGGAAGACCCCTCGGAGAACCTTGAAGACGACGACATTGCTCATTGCTTCCA 78981

Qy 61 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 92
Db 78980 GTATTACTTAGCCAGCCGGTGCTGGCTTTT 78949

RESULT 7
BA000007_42/c
WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA0000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000
BA000007_43	4300001	4410000
BA000007_44	4400001	4510000
BA000007_45	4500001	4610000
BA000007_46	4600001	4710000
BA000007_47	4700001	4810000
BA000007_48	4800001	4910000
BA000007_49	4900001	5010000
BA000007_50	5000001	5110000
BA000007_51	5100001	5210000
BA000007_52	5200001	5310000
BA000007_53	5300001	5410000
BA000007_54	5400001	5498450

Continuation (43 of 55) of BA000007 from base 4200001 (BA000007 Escherichia coli O157:H7

Query Match 100.0%; Score 92; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.3e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGGATCAGAGACCTCGCGAGAACCTGAAAGCAGACATTCGTCTCATTGCTTCCA 60
Db 100272 CCGGATCAGAGACCTCGCGAGAACCTGAAAGCAGACATTCGTCTCATTGCTTCCA 100213
Qy 61 GTATTACTTAGCCAGCGGGTCTGGCTTTT 92
Db 100212 GTATTACTTAGCCAGCGGGTCTGGCTTTT 100181

RESULT 8
BA000007_43/c
WPCOMMENT

Sequence split into 55 fragments LOCUS BA000007 Accession BA0000007

Fragment Name	Begin	End
BA000007_00	1	110000
BA000007_01	100001	210000
BA000007_02	200001	310000
BA000007_03	300001	410000
BA000007_04	400001	510000
BA000007_05	500001	610000
BA000007_06	600001	710000
BA000007_07	700001	810000
BA000007_08	800001	910000
BA000007_09	900001	1010000
BA000007_10	1000001	1110000
BA000007_11	1100001	1210000
BA000007_12	1200001	1310000
BA000007_13	1300001	1410000
BA000007_14	1400001	1510000
BA000007_15	1500001	1610000
BA000007_16	1600001	1710000
BA000007_17	1700001	1810000
BA000007_18	1800001	1910000
BA000007_19	1900001	2010000
BA000007_20	2000001	2110000
BA000007_21	2100001	2210000
BA000007_22	2200001	2310000
BA000007_23	2300001	2410000
BA000007_24	2400001	2510000
BA000007_25	2500001	2610000
BA000007_26	2600001	2710000
BA000007_27	2700001	2810000
BA000007_28	2800001	2910000
BA000007_29	2900001	3010000
BA000007_30	3000001	3110000
BA000007_31	3100001	3210000
BA000007_32	3200001	3310000
BA000007_33	3300001	3410000
BA000007_34	3400001	3510000
BA000007_35	3500001	3610000
BA000007_36	3600001	3710000
BA000007_37	3700001	3810000
BA000007_38	3800001	3910000
BA000007_39	3900001	4010000
BA000007_40	4000001	4110000
BA000007_41	4100001	4210000
BA000007_42	4200001	4310000
BA000007_43	4300001	4410000
BA000007_44	4400001	4510000
BA000007_45	4500001	4610000
BA000007_46	4600001	4710000
BA000007_47	4700001	4810000
BA000007_48	4800001	4910000
BA000007_49	4900001	5010000
BA000007_50	5000001	5110000
BA000007_51	5100001	5210000
BA000007_52	5200001	5310000
BA000007_53	5300001	5410000
BA000007_54	5400001	5498450

Continuation (44 of 55) of BA000007 from base 4300001 (BA000007 Escherichia coli O157:H7

Query Match 100.0%; Score 92; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 4.3e-23;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCGGATCAGAGACCTCGCGAGAACCTGAAAGCAGACATTCGTCTCATTGCTTCCA 60
Db 272 CCGGATCAGAGACCTCGCGAGAACCTGAAAGCAGACATTCGTCTCATTGCTTCCA 213
Qy 61 GTATTACTTAGCCAGCGGGTCTGGCTTTT 92

Db	212	GTATTACTAGCCAGCGGCTGCTGCTTTT	181	CDS	<pre> /locus_tag="S4181" 1335..2165 /genes="yiaF" /locus_tag="S4181" /notes="residues 1 to 276 of 276 are 94.56 pct identical to residues 1 to 276 from Escherichia coli K-12 : B3554" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAP19149.1" /db_xref="GI:30043429" /translation="MILPGLRRKGILOACPGLSLQSRQTRVCRVRCALFLGSRKMATG- KSCSRFAPLAALLMVVLSGCGDKEGDORKAFIDFLQNTVMRSGERLPTLTADQKKQ KGFPEVDYAILGYGSOQVQAOQMDSGLRPVVDSVNAIRVPODYVTQSGPLREMGSLGV LAQLQNAKLQADAAHSALKQSDDLKPFVDQAFKVTVPADALQPLIPAAQTFTQQL VMVDYIAQQGTQVSFVANGIQFTSQQASEYNKLIAPLPAQHQAQFNAQTAVTATQ " complement(2215..3189) /genes="yiaE" /locus_tag="S4182" complement(2215..3189) /genes="yiaE" /locus_tag="S4182" /function="putative enzyme; Not classified" /notes="residues 1 to 324 of 324 are 94.13 pct identical to residues 5 to 328 from Escherichia coli K-12 : B3553" /codon_start=1 /transl_table=11 /product="putative dehydrogenase" /protein_id="AAP19150.1" /db_xref="GI:30043430" /translation="MKPSVILYKALPDDLLQRLQAHFTVHQVANLSPQTVQNAIFA EAEGLLGSNENVDAALEKMPKLRTSTISVGDYDFVDALTARKILLMHTPTVLTET AQTDLMALVLTARRVVEVAERKAGEWTASIGPDYGTDVHHTKILGIVGMGRIGMAL AORVHFGFNNPILYNARRHHKEAEERFNARYCDLDTLLQESDFVCILPLTDTHILF GAQPAKMKSSAIFINAGRGPVVDENALIAALQKGIHAAGLDVFEQEPVLSVDSPLLS MANVAVPHIGSNATHETRYGMACAVDNLIDALQKVKNCVNPVHAD" complement(3293..3919) /genes="yiaD" /locus_tag="S4183" complement(3293..3919) /genes="yiaD" /locus_tag="S4183" /function="putative membrane; Not classified" /notes="residues 1 to 208 of 208 are 88.94 pct identical to residues 12 to 219 of 219 from Escherichia coli K-12 : B3552" /codon_start=1 /transl_table=11 /product="putative outer membrane protein" /protein_id="AAP19151.1" /db_xref="GI:30043431" /translation="MSGALAVSGCTTNPYTGEREAGKSAIGAGLSLVGAGIGALSSS KDRKGALIGAAAGALGGGYVMDQEAKLDRDKMRGTGVSVTRSGDNIILNMPNN VTFDSSAPLKPAGANTLPGVAMVLPKPTANNVIGYTDSTGCHDLNMLLSQQRADS VASALITQGVDSARIRTKGLGPANPIASNSTAEKQNRNRVEITLSPL" 4160..6439 /genes="biseC" /locus_tag="S4184" 4160..6439 /genes="biseC" /locus_tag="S4184" /function="enzyme; Biosynthesis of cofactors, carriers; Biotin" /notes="residues 21 to 759 of 759 are 98.78 pct identical to residues 1 to 739 of 739 from Escherichia coli K-12 : B3551" /codon_start=1 /transl_table=11 /product="biotin sulfoxide reductase" /protein_id="AAP19152.1" </pre>
RESULT 9	AE016992			gene	
LOCUS	AE016992	289816 bp	DNA	CDS	
DEFINITION	Shigella flexneri 2a str. 2457T	section 15 of 16 of the complete genome.			
ACCESSION	AE016992	AE014073			
VERSION	AE016992.1	GI:30043426			
KEYWORDS	Shigella flexneri 2a str. 2457T				
SOURCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
ORGANISM	Enterobacteriaceae; Shigella.				
REFERENCE	1 (bases 1 to 289816)				
AUTHORS	Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,				
	Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,				
	Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,				
	Schwartz, D.C. and Blattner, F.R.				
TITLE	Complete Genome Sequence and Comparative Genomics of Shigella				
JOURNAL	flexneri Serotype 2a Strain 2457T				
PUBLISHED	Infect. Immun. 71 (5), 2775-2786 (2003)				
REFERENCE	2 (bases 1 to 289816)				
AUTHORS	Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,				
	Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,				
	Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,				
	Schwartz, D.C. and Blattner, F.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (13-JUN-2002) Genetics Laboratory, University of				
FEATURES	Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA				
source	Location/Qualifiers				
	1..289816				
	/organism="Shigella flexneri 2a str. 2457T"				
	/mol_type="genomic DNA"				
	/strain="2457T"				
	/serotype="2a"				
	/db_xref="taxon:198215"				
	complement(238..450)				
gene	/gene="cspA"				
	/locus_tag="S4179"				
CDS	complement(238..450)				
	/gene="cspA"				
	/locus_tag="S4179"				
	/function="regulator; Adaptations, atypical conditions"				
	/notes="residues 1 to 70 of 70 are 100.00 pct identical to				
	residues 1 to 70 of 70 from Escherichia coli K-12 : B3556"				
	/codon_start=1				
	/transl_table=11				
	/product="cold shock protein 7.4, transcriptional				
	activator of hns"				
	/protein_id="AAP19147.1"				
	/db_xref="GI:30043427"				
	/translation="MSGKWTGIKWNADKGFITPDGSKDVFVHFSAIQNDGYS				
	LDGQKVSFTIESGAKGPAAGNTSL"				
	complement(896..1021)				
gene	/gene="yiaG"				
	/locus_tag="S4180"				
CDS	complement(896..1021)				
	/gene="yiaG"				
	/locus_tag="S4180"				
	/notes="residues 1 to 41 of 41 are 100.00 pct identical to				
	residues 1 to 41 of 96 from Escherichia coli K-12 : B3558"				
	/codon_start=1				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="AAP19148.1"				
	/db_xref="GI:30043428"				
	/translation="MEYKDPMBLLSLEIQVFKDETKITLTHRTTCTEIBQL"				
	1335..2165				
gene	/genes="yiaF"				

```
/db_xref="GI:30043432"
/translation="MLVETDGETVFSSRGALATGMENSLSQSAVRDQVHSNTRVRPFMV
RKGFLASPNQGIHQDQKRVFVSWDEALDLHQHKKRIREAYGPASIPAGSYGWSRN
GLYKASTLQGYMALAGYTGHLGDYSTGAAQAIMPYVVGSEVYQQOTSPWLVLEH
SDVVLMSANPLNTLKIANNASDEQSLFVSALRDSQKKLICIDPMRSSTVDFGDKM
ENVAHMDTNDALMTGLIAHTLVENGWHDFAFLARCTTGAVFASVALLGESDIAKNAE
WAAEICGVGAAKIRELALIFHQNTIMLMAGWQKQOQFGEQKHWMILVILNMLGQIGT
PGGGFLSHFANGGNPTFRSVALSMSQSLPGGCDVADKIPIVARIVEALENPGGAYQ
HNGMRHFDPIRIFIMWAGANFTHQDNTNRLIRAWQKPELVISCFWFAAAKHADIV
LPATTSFERNDLTMTGDYNSQHLVPMKQVPPRYEARNDVDFAEISERWEGGYARF
TEGKSQLOLETFYNVARQASQOVELPFAEFAOANOLIEMPENPDSERFIRPADF
CHDPLAHLKTSAGKLEIFSORIADYGPCDCHPMWLEPDSWQNGNAPEQLOVLSAH
PAHRLHSQNYLSRLRELYAVANREPVIHPDDAQARGITEGDMVRWNSRGQTLGAW
ISEG1KPGVICIHEGAPDLDTADGICKNGAVNVLTKDLPSSRLGNCAGNTLAWL
EYKNGPELTLTAFEPSPASS"
complement(6408..6848)
/genes="yiaC"
/locus_tag="S4185"
complement(6408..6848)
/genes="yiaC"
/locus_tag="S4185"
/notes="residues 1 to 146 of 146 are 97.94 pct identical to
residues 1 to 146 of 146 from Escherichia coli K-12 :
B3550"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP19153.1"
/db_xref="GI:30043433"
/translation="MIREAQRSELPALLELWLESTWGHFFIKANYWRDCIPLVRDAY
LANAQNWVEERGKLLGFSVINEGRFLAAMFVAPKAVRRGIGKALMQYVQQRYPHML
EYVQKNQPAIDFYAQGFHIVDCAWQDETQLPTWMSVPVQTL"
complement(6845..7408)
/genes="tag"
/locus_tag="S4186"
complement(6845..7408)
/genes="tag"
/locus_tag="S4186"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/notes="residues 1 to 187 of 187 are 97.32 pct identical to
residues 1 to 187 of 187 from Escherichia coli K-12 :
B3549"
/codon_start=1
/transl_table=11
/product="3-methyl-adenine DNA glycosylase I"
/protein_id="AAP19154.1"
/db_xref="GI:30043434"
/translation="MERCWVSGGLPIYAHDNWGVPTDSSKLFEMICFGQQAGL
SWITLKKRNYRAPHQDPVVKVAAQBEDVELRVQDAGIIRHRGKIQAIIGNARAY
LQMEQNGEPDFVMSFVNHQPVQVQTATLSIPTSTASDALSALKKRGKPFVGT
ICYSFPAQAGLVNDHVVGGCCYLGKNP"
7566..8264
/genes="yhjY"
/locus_tag="S4187"
7566..8264
/genes="yhjY"
/locus_tag="S4187"
/function="putative enzyme; Not classified"
/notes="residues 1 to 232 of 232 are 99.13 pct identical to
residues 1 to 232 of 232 from Escherichia coli K-12 :
B3548"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfL"
/protein_id="AAN82583.1"
/db_xref="GI:26110398"
/translation="WNKFIKVALVCAVLATLTACTGHENRDKNSYDYLHHPAISIS
KIIGCGGPTAQ"
complement(432..1517)
/genes="yhfS"
/locus_tag="c4146"
complement(432..1517)
/genes="yhfS"
/locus_tag="c4146"
complement(432..1517)
/notes="Escherichia coli K-12 ortholog: b3376"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfS"
/protein_id="AAN82584.1"
/db_xref="GI:26110399"
/translation="MKTFFLOSLTIEAQKQFALVDTICRHFPGCEFLTCDGLGUTP
GLNQPRITQRVEQVLADFAHQAAALVQAGTGAIRAAALALKSGRLLVHDAPVYP
TRVVIQEWGLTLITADFNDLISALKVVDEQPDAAVQHTRQQQDRYVLADVLATL
RAAGVPALTDNNYAVMKVARIQCEGANVSTFCFKLFGPEGVGA VGVADVTSIRA
TLYSGGSIQGAQALEVLRGLVAPVHVAQVGSERLLALLNGGAVKSAVIANA
KSKVLIVFHFQPIAARVLEBAKQKALPYFVGAESKYEIPFLYRLSGTFRQANPQLE
HCAIRINPRSGEETILRLRESIASI"
complement(1529..2833)
/genes="yhfT"
/locus_tag="c4147"
complement(1529..2833)
```

```
RESULT 10
LOCUS
DEFINITION
ACCESSION
VERSION
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS
gene
CDS
gene
CDS
gene
CDS
```

```
AE016768 301660 bp DNA linear BCT 24-DEC-2002
Escherichia coli CFT073 section 14 of 18 of the complete genome.
AE016768 AE014075
AE016768.1 GI:26110397
Escherichia coli CFT073
Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 301660)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
2 (bases 1 to 301660)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..301660
/molecule="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
157..324
/genes="yhfL"
/locus_tag="c4145"
157..324
/genes="yhfL"
/locus_tag="c4145"
/notes="Escherichia coli K-12 ortholog: b3369; Escherichia
coli O157:H7 ortholog: 24730"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfL"
/protein_id="AAN82583.1"
/db_xref="GI:26110398"
/translation="WNKFIKVALVCAVLATLTACTGHENRDKNSYDYLHHPAISIS
KIIGCGGPTAQ"
complement(432..1517)
/genes="yhfS"
/locus_tag="c4146"
complement(432..1517)
/genes="yhfS"
/locus_tag="c4146"
complement(432..1517)
/notes="Escherichia coli K-12 ortholog: b3376"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yhfS"
/protein_id="AAN82584.1"
/db_xref="GI:26110399"
/translation="MKTFFLOSLTIEAQKQFALVDTICRHFPGCEFLTCDGLGUTP
GLNQPRITQRVEQVLADFAHQAAALVQAGTGAIRAAALALKSGRLLVHDAPVYP
TRVVIQEWGLTLITADFNDLISALKVVDEQPDAAVQHTRQQQDRYVLADVLATL
RAAGVPALTDNNYAVMKVARIQCEGANVSTFCFKLFGPEGVGA VGVADVTSIRA
TLYSGGSIQGAQALEVLRGLVAPVHVAQVGSERLLALLNGGAVKSAVIANA
KSKVLIVFHFQPIAARVLEBAKQKALPYFVGAESKYEIPFLYRLSGTFRQANPQLE
HCAIRINPRSGEETILRLRESIASI"
complement(1529..2833)
/genes="yhfT"
/locus_tag="c4147"
complement(1529..2833)
```


WPCOMMENT

Sequence split into 48 fragments

LOCUS AE017220

Accession AE017220

Fragment Name

Begin

End

1

100001

210000

200001

310000

300001

410000

400001

510000

500001

610000

600001

710000

700001

810000

800001

910000

900001

1010000

1000001

1110000

1100001

1210000

1200001

1310000

1300001

1410000

1400001

1510000

1500001

1610000

1600001

1710000

1700001

1810000

1800001

1910000

1900001

2010000

2000001

2110000

2100001

2210000

2200001

2310000

2300001

2410000

2400001

2510000

2500001

2610000

2600001

2710000

2700001

2810000

2800001

2910000

2900001

3010000

3000001

3110000

3100001

3210000

3200001

3310000

3300001

3410000

3400001

3510000

3500001

3610000

3600001

3710000

3700001

3810000

3800001

3910000

3900001

4010000

4000001

4110000

4100001

4210000

4200001

4310000

4300001

4410000

4400001

4510000

4500001

4610000

4600001

4710000

4700001

4755700

Continuation (37 of 48) of AE017220 from base 3600001 (AE017220 Salmonella enterica subs

Query Match

Best Local Similarity

70.4%; Score 64.8; DB 1; Length 110000;

Matches

75; Conservative

0; Mismatches

17; Indels

0; Gaps

0;

QY

1

CCGGATCAGGAGACCTCGCGAGAACCTGAAAGCAGCAGCATTTGCTCACATTGCTTCCA

60

Db

68800

CCGGTTCAGGGGAAACCCCTACGAGNACTGAAAGCAGCAGCATTTGCTCACATTGCTTCCA

68741

QY

61

GTATTACTTAGCCAGCCGGTGCTGCTTTT

92

Db

68740

GTATTATTGGCCAGCTTTTGTGCTTTT

68709

RESULT 12

AE008863/c

LOCUS

Salmonella typhimurium LT2,

23506 bp

DNA

linear

BCT 09-AUG-2005

DEFINITION

Salmonella typhimurium LT2, section 167 of 220 of the complete genome.

ACCESSION

AE008863

AE006468

VERSION

AE008863.1

GI:16422092

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBMED

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

COMMENT

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.Pangeasystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

Location/Qualifiers

1..23506

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

complement(120..2581)

/gene="glpP"

/note="Synonym: STM3534"

complement(120..2567)

/gene="glpP"

/EC_number="2.4.1.1"

/note="similar to E. coli glycogen phosphorylase (AAC76453.1); Blastp hit to AAC76453.1 (815 aa), 93% identity in aa 1 - 815"

/codon_start=1

/transl_table=11

/product="glycogen phosphorylase"

/protein_id="AAL22394.1"

/db_xref="GI:16422093"

/translation="MNAPFTYASPTLSVEALKHSIAYKLMFTIGKDPVIANKHEWLNA TLFAVDRDLVERLWRSNQAQSQETROYIUSMEFLIGRTLSNALLSGIYDDVGKAL EAMGLDLDELDENDPGLNGGLGRLAACFLDSLATLGLPGRGYGYDGMFKQNI VDGROKESPDYWLVEKPNFWRKHNRKTVLFGGRIQOEKKARWIEEILAVYDQ IIPGYDTATNTLRNNAQASSEINLGNKQDGYFAAEDKNHSENVSRVLPDSTY SGRELRLQEFPLVSATVQDILHRHYQLHKTYENLADKTAIHLNDTHPVLSPELMRL

CP000026_05 500001 610000
CP000026_06 600001 710000
CP000026_07 700001 810000
CP000026_08 800001 910000
CP000026_09 900001 1010000
CP000026_10 1000001 1110000
CP000026_11 1100001 1210000
CP000026_12 1200001 1310000
CP000026_13 1300001 1410000
CP000026_14 1400001 1510000
CP000026_15 1500001 1610000
CP000026_16 1600001 1710000
CP000026_17 1700001 1810000
CP000026_18 1800001 1910000
CP000026_19 1900001 2010000
CP000026_20 2000001 2110000
CP000026_21 2100001 2210000
CP000026_22 2200001 2310000
CP000026_23 2300001 2410000
CP000026_24 2400001 2510000
CP000026_25 2500001 2610000
CP000026_26 2600001 2710000
CP000026_27 2700001 2810000
CP000026_28 2800001 2910000
CP000026_29 2900001 3010000
CP000026_30 3000001 3110000
CP000026_31 3100001 3210000
CP000026_32 3200001 3310000
CP000026_33 3300001 3410000
CP000026_34 3400001 3510000
CP000026_35 3500001 3610000
CP000026_36 3600001 3710000
CP000026_37 3700001 3810000
CP000026_38 3800001 3910000
CP000026_39 3900001 4010000
CP000026_40 4000001 4110000
CP000026_41 4100001 4210000
CP000026_42 4200001 4310000
CP000026_43 4300001 4410000
CP000026_44 4400001 4510000
CP000026_45 4500001 4595229

Continuation (35 of 46) of CP000026 from base 3400001 (CP000026 Salmonella enterica subsp. enterica)

Query Match 69.3%; Score 63.8; DB 1; Length 110000;
Best Local Similarity 81.3%; Pred. No. 2.le-12;
Matches 74; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 GCGATCAGGAGACCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATGCTTCCAG 61
DB 106615 GCATTAGGGGAAACCCCTACGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATGCTTCCAG 106556

QY 62 TATTACTTAGCCAGCGGGTGTGGCTTTT 92
DB 106555 TATTATTGGCAGCTTTTGTGGCTTTT 106525

RESULT 14
CP000026_35/C
WPCOMMENT

Sequence split into 46 fragments LOCUS CP000026 Accession CP0000026

Fragment Name	Begin	End
CP000026_00	1	110000
CP000026_01	100001	210000
CP000026_02	200001	310000
CP000026_03	300001	410000
CP000026_04	400001	510000
CP000026_05	500001	610000
CP000026_06	600001	710000
CP000026_07	700001	810000
CP000026_08	800001	910000
CP000026_09	900001	1010000
CP000026_10	1000001	1110000
CP000026_11	1100001	1210000

1200001 1310000
1300001 1410000
1400001 1510000
1500001 1610000
1600001 1710000
1700001 1810000
1800001 1910000
1900001 2010000
2000001 2110000
2100001 2210000
2200001 2310000
2300001 2410000
2400001 2510000
2500001 2610000
2600001 2710000
2700001 2810000
2800001 2910000
2900001 3010000
3000001 3110000
3100001 3210000
3200001 3310000
3300001 3410000
3400001 3510000
3500001 3610000
3600001 3710000
3700001 3810000
3800001 3910000
3900001 4010000
4000001 4110000
4100001 4210000
4200001 4310000
4300001 4410000
4400001 4510000
4500001 4595229

Continuation (36 of 46) of CP000026 from base 3500001 (CP000026 Salmonella enterica subsp. enterica)

Query Match 69.3%; Score 63.8; DB 1; Length 110000;
Best Local Similarity 81.3%; Pred. No. 2.le-12;
Matches 74; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 2 GCGATCAGGAGACCTCGCGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATGCTTCCAG 61
DB 6615 GCATTAGGGGAAACCCCTACGGAGAACCTGAAAGCAGCAGCATTTGCTCACAATGCTTCCAG 6556

QY 62 TATTACTTAGCCAGCGGGTGTGGCTTTT 92
DB 6555 TATTATTGGCAGCTTTTGTGGCTTTT 6525

RESULT 15
AL627281
LOCUS

DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18, complete chromosome, segment 17/20.

ACCESSION AL627281 AL513382

VERSION AL627281.1 GI:16504930

KEYWORDS

SOURCE Salmonella enterica subsp. enterica serovar Typhi

ORGANISM Salmonella enterica subsp. enterica serovar Typhi Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

REFERENCE 1 (bases 1 to 285050)
Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T., Sebaihia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltwell, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagers, K., Krogh, A., Larsen, T.S., Leather, S., Mout, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrell, B.G.
Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18

TITLE

JOURNAL Nature 413 (6858), 848-852 (2001)


```
/gene="STY4115"
/notes="Pfam match to entry PF00480 ROK, ROK family, score
122.90, E-value 1.8e-35"
complement(5424..5427)
/notes="possible RBS"
5573..5579
/notes="possible RBS"
5586..7124
/genes="STY4116"
/notes="synonym: aldB"
5586..7124
/genes="STY4116"
/EC_number="1.2.1.22"
/notes="Fasta hit to ALDA_ECOLI (478 aa), 33% identity in
478 aa overlap
Fasta hit to FEAB_ECOLI (499 aa), 34% identity in 481 aa
overlap
Fasta hit to YNEI_ECOLI (462 aa), 31% identity in 465 aa
overlap
Fasta hit to DHAB_ECOLI (489 aa), 38% identity in 478 aa
overlap
Fasta hit to DHAL_ECOLI (495 aa), 40% identity in 481 aa
overlap
Fasta hit to YDCW_ECOLI (474 aa), 35% identity in 481 aa
overlap
Fasta hit to GABD_ECOLI (482 aa), 34% identity in 485 aa
overlap
Orthologue of E. coli aldB (ALDB_ECOLI); Fasta hit to
ALDB_ECOLI (512 aa), 96% identity in 512 aa overlap"
/codon_start=1
/translation=11
/product="aldenhyde dehydrogenase B"
```

```
Query Match      64.1%; Score 59; DB 1; Length 265050;
Best Local Similarity 78.0%; Pred. No. 1.5e-10;
Matches 71; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      2  GCGATCAGGAAGACCTCGCGGAGAACCTGAAGACGACGACATTGCTCACATTGCTTCCAG 61
Db      157180  GCATTGAGGGGACCCCTACGGATACCTGAAGACGACATTGCTCACATTGCTTCCAG 157239

QY      62  TATTACTTAGCCAGCGGGTCTGGCTTTT 92
Db      157240  TTTTATTTGGCAGCTTTTGTGCGCTTTT 157270
```

Search completed: February 16, 2006, 05:50:26
Job time : 1806 secs